1600

RAW SEQUENCE LISTING
PATENT APPLICATION: **US/09/470,667**

DATE: 02/26/2003
TIME: 13:27:00

Input Set : **N:\AMC\US09470667.raw**
Output Set: **N:\CRF4\02262003\I470667.raw**

```
 1 <110> APPLICANT: Asakura, Akira
 2         Hoshino, Tatsuo
 3         Ojima, Setsuko
 4         Shinjoh, Masako
 5         Tomiyama, Noribumi
 6 <120> TITLE OF INVENTION: Novel Alcohol/Aldehyde Dehydrogenases
 7 <130> FILE REFERENCE: C38435/109700CON
 8 <140> CURRENT APPLICATION NUMBER: US/09/470,667
 9 <141> CURRENT FILING DATE: 1999-12-22
10 <150> PRIOR APPLICATION NUMBER: US 08/934,506
11 <151> PRIOR FILING DATE: 1997-09-19
12 <160> NUMBER OF SEQ ID NOS: 12
13 <170> SOFTWARE: PatentIn version 3.1
15 <210> SEQ ID NO: 1
16 <211> LENGTH: 1740
17 <212> TYPE: DNA
18 <213> ORGANISM: Gluconobacter oxydans
19 <400> SEQUENCE: 1
20       atgaaaccga cttcgctgct ttgggccagt gctggcgcac ttgcattgct tgccgcaccc    60
21       gcctttgctc aagtgacccc cgtcaccgat gaattgctgg cgaacccgcc cgctggtgaa   120
22       tggatcagct acggtcagaa ccaagaaaac taccgtcact cgcccctgac gcagatcacg   180
23       actgagaacg tcggccaact gcaactggtc tgggcgcgcg gcatgcagcc gggcaaagtc   240
24       caagtcacgc ccctgatcca tgacggcgtc atgtatctgg caaacccggg cgacgtgatc   300
25       caggccatcg acgccaaaac tggcgatctg atctgggaac accgccgcca actgccgaac   360
26       atcgccacgc tgaacagctt tggcgagccg acccgcggca tggcgctgta cggcaccaac   420
27       gtttactttg tttcgtggga caaccacctg gtcgccctcg acaccgcaac tggccaagtg   480
28       acgttcgacg tcgaccgcgg ccaaggcgaa gacatggttt cgaactcgtc gggcccgatc   540
29       gtggcaaacg gcgtgatcgt tgccggttcg acctgccaat actcgccgtt cggctgcttt   600
30       gtctcgggcc acgactcggc caccggtgaa gagctgtggc gcaactactt catcccgcgc   660
31       gctggcgaag agggtgatga gacttggggc aacgattacg aagcccgttg gatgaccggt   720
32       gcctgggggcc agatcaccta tgaccccgtc accaaccttg tccactacgg ctcgaccgct   780
33       gtgggtccgg cgtcggaaac ccaacgcggc accccgggcg gcacgctgta cggcacgaac   840
34       acccgtttcg ccgtgcgtcc tgacacgggc gagattgtct ggcgtcacca gaccctgccc   900
35       cgcgacaact gggaccagga atgcacgttc gagatgatgg tcaccaatgt ggatgtccaa   960
36       ccctcgaccg agatggaagg tctgcagtcg atcaacccga acgccgcaac tggcgagcgt  1020
37       cgcgtgctga ccggcgttcc gtgcaaaacc ggcaccatgt ggcagttcga cgccgaaacc  1080
38       ggcgaattcc tgtgggcccg tgataccaac taccagaaca tgatcgaatc catcgacgaa  1140
39       aacggcatcg tgaccgtgaa cgaagatgcg atcctgaagg aactggatgt tgaatatgac  1200
40       gtctgcccga ccttcttggg cggccgcgac tggccgtcgg ccgcactgaa ccccgacagc  1260
41       ggcatctact tcatcccgct gaacaacgtc tgctatgaca tgatggccgt cgatcaggaa  1320
42       ttcacctcga tggacgtcta taacaccagc aacgtgacca agctgccgcc cggcaaggat  1380
43       atgatcggtc gtattgacgc gatcgacatc agcacgggtc gtacgctgtg gtcggtcgaa  1440
44       cgtgctgcgg cgaactattc gcccgtcttg tcgaccggcg gcggcgttct gttcaacggt  1500
```

RAW SEQUENCE LISTING
PATENT APPLICATION: **US/09/470,667**

DATE: 02/26/2003
TIME: 13:27:00

Input Set : **N:\AMC\US09470667.raw**
Output Set: **N:\CRF4\02262003\I470667.raw**

```
45        ggtacggatc gttacttccg cgccctcagc caagaaaccg gcgagaccct gtggcagacc    1560
46        cgccttgcaa ccgtcgcgtc gggccaggcc atctcttacg aggttgacgg catgcaatat    1620
47        gtcgccatcg caggtggtgg tgtcagctat ggctcgggcc tgaactcggc actggctggc    1680
48        gagcgagtcg actcgaccgc catcggtaac gccgtctacg tcttcgccct gccgcaataa    1740
50  <210> SEQ ID NO: 2
51  <211> LENGTH: 1740
52  <212> TYPE: DNA
53  <213> ORGANISM: Gluconobacter oxydans
54  <400> SEQUENCE: 2
55        atgaagacgt cgtctttgct ggttgcgagc gttgccgcgc ttgcaagcta tagctccttt      60
56        gcgcttgctc aagtgacccc cgtcaccgat gaattgctgg cgaacccgcc cgctggtgaa     120
57        tggatcagct acggtcagaa ccaagaaaac taccgtcact cgcccctgac gcagatcacg     180
58        actgagaacg tcggccaact gcaactggtc tgggcgcgcg gcatgcagcc gggcaaagtc     240
59        caagtcacgc ccctgatcca tgacggcgtc atgtatctgg caaacccggg cgacgtgatc     300
60        caggccatcg acgccaaaac tggcgatctg atctgggaac accgccgcca actgccgaac     360
61        atcgccacgc tgaacagctt tggcgagccg acccgcggca tggcgctgta cggcaccaac     420
62        gtttactttg tttcgtggga caaccacctg gtcgccctcg acaccgcaac tggccaagtg     480
63        acgttcgacg tcgaccgcgg ccaaggcgaa gacatggttt cgaactcgtc gggcccgatc     540
64        gtggcaaacg gcgtgatcgt tgccggttcg acctgccaat actcgccgtt cggctgcttt     600
65        gtctcgggcc acgactcggc caccggtgaa gagctgtggc gcaactactt catcccgcgc     660
66        gctggcgaag agggtgatga gacttggggc aacgattacg aagcccgttg gatgaccggc     720
67        gtctggggtc agatcaccta tgaccccgtt ggcggccttg tccactacgg ctcgtcggct     780
68        gttggcccgg cttcggaaac ccagcgcggc accaccggcg gcaccatgta cggcaccaac     840
69        acccgtttcg ctgtccgtcc cgagactggc gagatcgtct ggcgtcacca aactctgccc     900
70        cgcgacaact gggaccaaga gtgcaccttc gagatgatgg ttgccaacgt tgacgtgcag     960
71        cccgcagctg acatggacgg cgtccgctcg atcaacccga acgccgccac cggcgagcgt    1020
72        cgcgttctga ccggcgttcc gtgcaaaacc ggcaccatgt ggcagttcga cgccgaaacc    1080
73        ggcgaattcc tgtgggcccg tgacaccagc tacgagaaca tcatcgaatc gatcgacgaa    1140
74        aacggcatcg tgaccgtcga cgagtcgaaa gttctgaccg agctggacac cccctatgac    1200
75        gtctgcccgc tgctgctggg tggccgtgac tggccgtcgg ctgcgctgaa ccccgatacc    1260
76        ggcatctact ttatcccgct gaacaacacc tgcatggata tcgaagctgt cgaccaggaa    1320
77        ttcagctcgc tggacgtgta caaccaaagc ctgaccgcca aaatggcacc gggtaaagag    1380
78        ctggttggcc gtatcgacgc catcgacatc agcacaggcc gcaccctgtg gaccgctgag    1440
79        cgcgaagcct cgaactacgc gcctgtcctg tcgaccgctg gcggcgttct gttcaacggc    1500
80        ggcaccgacc gttacttccg cgctctcagc caagagaccg gcgagaccct gtggcagacc    1560
81        cgtctggcga ctgtcgcttc gggccaagct gtctcgtacg agatcgacgg cgtccaatac    1620
82        atcgccatcg gcggcggcgg cacgacctat ggttcgttcc acaaccgtcc cctggccgag    1680
83        ccggtcgact cgaccgcgat cggtaatgcg atgtacgtct tcgcgctgcc ccagcaataa    1740
85  <210> SEQ ID NO: 3
86  <211> LENGTH: 1737
87  <212> TYPE: DNA
88  <213> ORGANISM: Gluconobacter oxydans
89  <400> SEQUENCE: 3
90        atgaaactga cgaccctgct gcaaagcagc gccgccctgc ttgtgcttgg caccattccc      60
91        gcccttgccc aaaccgccat caccgatgaa atgctggcga acccgcccgc tggtgaatgg     120
92        atcaactacg gtcagaacca agagaactac cgccactcgc ccctgacgca gattaccgca     180
93        gacaacgtcg gccaactgca actggtctgg gcgcgcggta tggaagcggg caagatccaa     240
94        gtgaccccgc ttgtccatga cggcgtcatg tatctggcaa accccggtga cgtgatccag     300
95        gccatcgacg ccgcgaccgg cgatctgatc tgggaacacc gccgccaact gccgaacatc     360
```

```
 96        gccacgctga acagctttgg tgagccgacc cgcggcatgg ccctctatgg caccaacgtc    420
 97        tatttcgtct cgtgggacaa ccacttggtc gcgctggaca cctcgaccgg ccaagtcgta    480
 98        ttcgacgtcg atcgcggtca aggcacggat atggtctcga actcgtccgg cccgattgtc    540
 99        gccaatggcg tcatcgttgc gggctcgacc tgtcagtatt cgccgttcgg ctgtttcgtt    600
100        tcgggccacg actcggccac cggtgaagag ctgtggcgca acaccttat cccgcgcgcc    660
101        ggcgaagagg gtgatgagac ctggggcaat gattacgagg cccgctggat gaccggcgtt    720
102        tggggccaga tcacctatga ccccgttggc ggccttgtcc actacggcac ctcagcagtt    780
103        ggccctgcgg ccgagattca gcgcggcacc gttggcggct cgatgtatgg caccaacacc    840
104        cgctttgctg tccgccccga gaccggcgag atcgtctggc gtcaccaaac tctgccccgc    900
105        gacaactggg accaagagtg tacgttcgag atgatggtcg tcaacgtcga cgtccagccc    960
106        tcggctgaga tggaaggcct gcacgccatc aaccccgatg ccgccacggg cgagcgtcgc   1020
107        gttgtgaccg gcgttccgtg caagaacggc accatgtggc agttcgacgc cgaaaccggc   1080
108        gaattcctgt gggcgcgcga caccagctat cagaacctga tcgaaagcgt cgatcccgat   1140
109        ggtctggtgc atgtgaacga agatctggtc gtgaccgagc tggaagtggc ctatgaaatc   1200
110        tgcccgacct tcctgggtgg ccgcgactgg ccgtcggctg cgctgaaccc cgatactggc   1260
111        atctatttca tcccgctgaa caacgcctgt agcggtatga cggctgtcga ccaagagttc   1320
112        agctcgctcg atgtgtataa cgtcagcctc gactataaac tgtcgcccgg ttcggaaaac   1380
113        atgggccgta tcgacgccat cgacatcagc accggccgca cgctgtggtc ggctgaacgc   1440
114        tacgcctcga actacgcgcc tgtcctgtcc accggcggcg gcgtgctgtt caacggcggc   1500
115        accgaccgtt acttccgcgc cctcagccaa gagaccggcg agacgctgtg gcagacccgt   1560
116        ctggcgactg tcgcctcggg tcaagcgatt tcctatgaga tcgacggcgt gcaatatgtc   1620
117        gccatcgggc gcggcggcac cagctatggc agcaaccaca accgcgccct gaccgagcgg   1680
118        atcgactcga ccgccatcgg cagcgcgatc tatgtctttg ctctgccgca gcagtaa       1737
120  <210> SEQ ID NO: 4
121  <211> LENGTH: 1740
122  <212> TYPE: DNA
123  <213> ORGANISM: Gluconobacter oxydans
124  <400> SEQUENCE: 4
125        atgaacccca caacgctgct tcgcaccagc gcggccgtgc tattgcttac cgcgcccgcc     60
126        gcattcgcgc aggtaacccc gattaccgat gaactgctgg cgaacccgcc cgctggtgaa    120
127        tggattaact acggccgcaa ccaagaaaac tatcgccact cgcccctgac ccagatcact    180
128        gccgacaacg ttggtcagtt gcaactggtc tgggcccgcg ggatggaggc ggggccgta    240
129        caggtcacgc cgatgatcca tgatggcgtg atgtatctgg caaaccccgg tgatgtgatc    300
130        caggcgctgg atgcgcaaac aggcgatctg atctgggaac accgccgcca actgcccgcc    360
131        gtcgccacgc taaacgcccc aggcgaccgc aagcgcggcg tcgccctta cggcacgagc    420
132        ctctatttca gctcatggga caaccatctg atcgcgctgg atatggagac gggccaggtc    480
133        gtattcgatg tcgaacgtgg atcgggcgaa gacggcttga ccagtaacac cacggggccg    540
134        attgtcgcca atggcgtcat cgtcgcgggt tccacctgcc aatattcgcc ctatggatgc    600
135        tttatctcgg ggcacgattc cgcgacgggt gaggagctgt ggcgcaacca ctttatcccg    660
136        cagccgggcg aagagggtga cgagacttgg ggcaatgatt cgaggcgcg ctggatgacc    720
137        ggcgtctggg gtcagatcac ctatgatccc gtgacgaacc ttgtgttcta tggctcgacc    780
138        ggcgtgggcc cagcgtccga aacccagcgc ggcacgccgg cggcacgct gtatggcacc    840
139        aacacccgct ttgcggtgcg tcccgacacg ggcgagattg tctggcgtca ccagaccctg    900
140        ccgcgcgaca actgggacca agaatgcacg ttcgagatga tggtcgccaa cgtcgatgtg    960
141        caaccctcgg ccgagatgga gggtctgcgc gccatcaacc ccaatgcggc gacgggcgag   1020
142        cgccgtgtgc tgacgggtgc gccttgcaag accggcacga tgtggtcgtt tgatgcggcc   1080
143        tcgggcgaat tcctgtgggc gcgtgatacc aactacacca atatgatcgc ctcgatcgac   1140
144        gagaccggcc ttgtgacggt gaacgaggat gcggtgctga aagagctgga cgttgaatat   1200
145        gacgtctgcc cgaccttcct gggtgggcgc gactggtcgt cagccgcact gaacccggac   1260
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/470,667

DATE: 02/26/2003
TIME: 13:27:00

Input Set : N:\AMC\US09470667.raw
Output Set: N:\CRF4\02262003\I470667.raw

```
146        accggcattt acttcttgcc gctgaacaat gcctgctacg atattatggc cgttgatcaa   1320
147        gagtttagcg cgctcgacgt ctataacacc agcgcgaccg caaaactcgc gccgggcttt   1380
148        gaaaatatgg gccgcatcga cgcgattgat atcagcaccg ggcgcacctt gtggtcggcg   1440
149        gagcgccctg cggcgaacta ctcgcccgtt ttgtcgacgg caggcggtgt ggtgttcaac   1500
150        ggcgggaccg accgctattt ccgtgccctc agccaggaaa ccggcgagac tttgtggcag   1560
151        gcccgtcttg cgacggtcgc gacggggcag gcgatcagct acgagttgga cggcgtgcaa   1620
152        tatatcgcca tcggtgcggg cggtctgacc tatggcacgc aattgaacgc gccgctggcc   1680
153        gaggcaatcg attcgacctc ggtcggtaat gcgatctatg tctttgcact gccgcagtaa   1740
155  <210> SEQ ID NO: 5
156  <211> LENGTH: 579
157  <212> TYPE: PRT
158  <213> ORGANISM: Gluconobacter oxydans
159  <220> FEATURE:
160  <221> NAME/KEY: SIGNAL
161  <222> LOCATION: (1)..(23)
162  <223> OTHER INFORMATION:
```

W--> 163 <400> 5

```
164        Met Lys Pro Thr Ser Leu Leu Trp Ala Ser Ala Gly Ala Leu Ala Leu
165        1               5                   10                  15
166        Leu Ala Ala Pro Ala Phe Ala Gln Val Thr Pro Val Thr Asp Glu Leu
167                    20                  25                  30
168        Leu Ala Asn Pro Pro Ala Gly Glu Trp Ile Ser Tyr Gly Gln Asn Gln
169                    35                  40                  45
170        Glu Asn Tyr Arg His Ser Pro Leu Thr Gln Ile Thr Thr Glu Asn Val
171            50                  55                  60
172        Gly Gln Leu Gln Leu Val Trp Ala Arg Gly Met Gln Pro Gly Lys Val
173        65                  70                  75                  80
174        Gln Val Thr Pro Leu Ile His Asp Gly Val Met Tyr Leu Ala Asn Pro
175                        85                  90                  95
176        Gly Asp Val Ile Gln Ala Ile Asp Ala Lys Thr Gly Asp Leu Ile Trp
177                    100                 105                 110
178        Glu His Arg Arg Gln Leu Pro Asn Ile Ala Thr Leu Asn Ser Phe Gly
179                    115                 120                 125
180        Glu Pro Thr Arg Gly Met Ala Leu Tyr Gly Thr Asn Val Tyr Phe Val
181            130                 135                 140
182        Ser Trp Asp Asn His Leu Val Ala Leu Asp Thr Ala Thr Gly Gln Val
183        145                 150                 155                 160
184        Thr Phe Asp Val Asp Arg Gly Gln Gly Glu Asp Met Val Ser Asn Ser
185                        165                 170                 175
186        Ser Gly Pro Ile Val Ala Asn Gly Val Ile Val Ala Gly Ser Thr Cys
187                    180                 185                 190
188        Gln Tyr Ser Pro Phe Gly Cys Phe Val Ser Gly His Asp Ser Ala Thr
189                    195                 200                 205
190        Gly Glu Glu Leu Trp Arg Asn Tyr Phe Ile Pro Arg Ala Gly Glu Glu
191            210                 215                 220
192        Gly Asp Glu Thr Trp Gly Asn Asp Tyr Glu Ala Arg Trp Met Thr Gly
193        225                 230                 235                 240
194        Ala Trp Gly Gln Ile Thr Tyr Asp Pro Val Thr Asn Leu Val His Tyr
195                        245                 250                 255
```

RAW SEQUENCE LISTING
PATENT APPLICATION: **US/09/470,667**

DATE: 02/26/2003
TIME: 13:27:00

Input Set : **N:\AMC\US09470667.raw**
Output Set: **N:\CRF4\02262003\I470667.raw**

```
196        Gly Ser Thr Ala Val Gly Pro Ala Ser Glu Thr Gln Arg Gly Thr Pro
197                    260             265             270
198        Gly Gly Thr Leu Tyr Gly Thr Asn Thr Arg Phe Ala Val Arg Pro Asp
199                    275             280             285
200        Thr Gly Glu Ile Val Trp Arg His Gln Thr Leu Pro Arg Asp Asn Trp
201                    290             295             300
202        Asp Gln Glu Cys Thr Phe Glu Met Met Val Thr Asn Val Asp Val Gln
203        305             310             315             320
204        Pro Ser Thr Glu Met Glu Gly Leu Gln Ser Ile Asn Pro Asn Ala Ala
205                    325             330             335
206        Thr Gly Glu Arg Arg Val Leu Thr Gly Val Pro Cys Lys Thr Gly Thr
207                    340             345             350
208        Met Trp Gln Phe Asp Ala Glu Thr Gly Glu Phe Leu Trp Ala Arg Asp
209                    355             360             365
210        Thr Asn Tyr Gln Asn Met Ile Glu Ser Ile Asp Glu Asn Gly Ile Val
211                    370             375             380
212        Thr Val Asn Glu Asp Ala Ile Leu Lys Glu Leu Asp Val Glu Tyr Asp
213        385             390             395             400
214        Val Cys Pro Thr Phe Leu Gly Gly Arg Asp Trp Pro Ser Ala Ala Leu
215                    405             410             415
216        Asn Pro Asp Ser Gly Ile Tyr Phe Ile Pro Leu Asn Asn Val Cys Tyr
217                    420             425             430
218        Asp Met Met Ala Val Asp Gln Glu Phe Thr Ser Met Asp Val Tyr Asn
219                    435             440             445
220        Thr Ser Asn Val Thr Lys Leu Pro Pro Gly Lys Asp Met Ile Gly Arg
221        450             455             460
222        Ile Asp Ala Ile Asp Ile Ser Thr Gly Arg Thr Leu Trp Ser Val Glu
223        465             470             475             480
224        Arg Ala Ala Ala Asn Tyr Ser Pro Val Leu Ser Thr Gly Gly Gly Val
225                    485             490             495
226        Leu Phe Asn Gly Gly Thr Asp Arg Tyr Phe Arg Ala Leu Ser Gln Glu
227                    500             505             510
228        Thr Gly Glu Thr Leu Trp Gln Thr Arg Leu Ala Thr Val Ala Ser Gly
229                    515             520             525
230        Gln Ala Ile Ser Tyr Glu Val Asp Gly Met Gln Tyr Val Ala Ile Ala
231        530             535             540
232        Gly Gly Gly Val Ser Tyr Gly Ser Gly Leu Asn Ser Ala Leu Ala Gly
233        545             550             555             560
234        Glu Arg Val Asp Ser Thr Ala Ile Gly Asn Ala Val Tyr Val Phe Ala
235                    565             570             575
236        Leu Pro Gln
238 <210> SEQ ID NO: 6
239 <211> LENGTH: 579
240 <212> TYPE: PRT
241 <213> ORGANISM: Gluconobacter oxydans
242 <220> FEATURE:
243 <221> NAME/KEY: SIGNAL
244 <222> LOCATION: (1)..(23)
245 <223> OTHER INFORMATION:
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/09/470,667**

DATE: 02/26/2003
TIME: 13:27:01

Input Set : **N:\AMC\US09470667.raw**
Output Set: **N:\CRF4\02262003\I470667.raw**

```
L:163 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:0
L:246 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:6,Line#:0
L:329 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:7,Line#:0
L:412 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:8,Line#:0
L:511 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:11,Line#:0
```